

found at major text locations. The paper concludes with a consideration of pedagogic implications, pointing up the value of information on pauses for the identification of writer difficulties or uncertainties.

Language testing

00-659 Breiner-Sanders, Karen E. (Georgetown U., Middlebury Coll., USA), **Lowe, Jr., Pardee** (U.S. Dept. of Defense), **Miles, John** (Ed. Testing Service, USA) and **Swender, Elvira** (ACTFL, NY, USA; *Email*: actflhq@aol.com). ACTFL Proficiency Guidelines–Speaking: revised 1999. *Foreign Language Annals* (New York, USA), **33**, 1 (2000), 13–18.

The ACTFL (American Council on the Teaching of Foreign Languages) Proficiency Guidelines–Speaking (1986) have gained widespread application as a metric against which to measure learners' functional competence; that is, their ability to accomplish linguistic tasks representing a variety of levels. Based on years of experience with oral testing in governmental institutions and on the descriptions of language proficiency used by Interagency Language Roundtable (ILR), the ACTFL Guidelines were an adaptation intended for use in academia (college and university levels particularly) in the United States. This article presents the recent revision of the Guidelines. The purposes of the revision are to make the document more accessible to those who have not received recent training in ACTFL oral proficiency testing, to clarify the issues that have divided testers and teachers, and to provide a corrective to what the committee perceived to have been possible misinterpretations of the descriptions provided in earlier versions of the Guidelines. The revision is presented as an additional step towards more adequately describing speaking proficiency.

00-660 Charnock, Ross (Université de Rouen, France). Le 'washback' et les tests intégrés. A propos de l'utilité pédagogique de l'évaluation linguistique. ['Washback' and integrated tests: the pedagogical usefulness of linguistic evaluation.] *ASp (Anglais de Spécialité)* (Bordeaux, France), **23/26** (1999), 391–416.

This paper begins from the premise that the testing methods traditionally used for the evaluation of the linguistic competence of students specialised in other disciplines can be shown to be unsatisfactory from various points of view. Tests based on discrete point items have low validity, while exercises involving free expression have poor reliability. It is argued, however, that integrated tests, based on continuous discourse, such as the cloze procedure, the C-test, proof-reading, dictation, and (possibly) reading aloud, seem to combine the advantages of both types of test, while avoiding the related drawbacks. In addition, they seem to be more useful from a pedagogical point of view, the favourable

washback effect thereby justifying their use as classroom exercises.

00-661 Clapham, Caroline (Lancaster U., UK). Assessment and testing. *Annual Review of Applied Linguistics* (New York, USA), **20** (2000), 147–61.

This article discusses the relationship between language testing and the other sub-disciplines of applied linguistics and also the relationship, as the author sees it, between testing and assessment. The article starts with a brief exploration of the term 'applied linguistics' and then goes on to discuss the role of language testing within this discipline, the relationship between testing and teaching, and the relationship between testing and assessment. The second part of the article mentions some areas of current concern to testers and discusses in more detail recent advances in the areas of performance testing, alternative assessment, and computer assessment. One of the aims of the article is to argue that the skills involved in language testing are necessary not only for those constructing all kinds of language proficiency assessments, but also for those other applied linguists who use tests or other elicitation techniques to help them gather language data for research. The article is followed by both an annotated and unannotated bibliography.

00-662 Edelenbos, Peter (U. of Groningen / The Netherlands Language Academy; *Email*: edelenbo@ppsw.rug.nl) and **Vinje, Marja P.** The assessment of a foreign language at the end of primary (elementary) education. *Language Testing* (London, UK), **17**, 2 (2000), 144–62.

This article examines assessment at the end of the primary phase with specific reference to levels of English. Within the framework of the Dutch national assessment scheme (PPON—the National Assessment Programme in Education), it describes the development and administration of tests to assess levels of learner attainment in listening, reading and word knowledge in English. The researchers examine a range of 'other' variables which might have influenced performance in listening at the end of primary school. These include pupil influences—e.g., socio-economic status and gender, and instructional influences—e.g., level of contact hours and pedagogic orientation of the learning programme itself. They also offer a comparative analysis of differences in learner performance in the same skill areas from the PPON 1991 and 1996 assessments. The research shows wide differences in performance between pupils from higher and lower social levels, with boys also outperforming girls. Time also emerges as a rather stable and significant variable. The assessments are seen as providing important insights into the possibilities and limitations of foreign language teaching and learning.

00-663 Gamaroff, Raphael (U. of Fort Hare, South Africa; *Email*: raphgam@border.co.za). Rater reliability in language assessment: the bug of all bears. *System* (Oxford, UK), **28**, 1 (2000), 31–53.

A major problem in essay assessment is how to achieve an overall reliable score based on the judgements of specific criteria such as topic relevance and grammatical accuracy. To investigate this question, the present author conducted a workshop on interrater reliability at a conference of the National Association of Educators of Teachers of English (NAETE) in South Africa where a group of 24 experienced educators of teachers of English were asked to assess two Grade 7 English essay protocols. The results revealed substantial variability in attention paid by raters to different criteria varying from penalising students for spelling and/or grammatical errors to glossing over these criteria and considering mainly content. To overcome the problem of rater variability some researchers recommend that more than one rater be used. The problem is that in the teaching situation there is rarely more than one rater available, who is usually the teacher of the subject. The advantages and disadvantages of using a single rater and more than one rater are examined. It is concluded that, whether one uses one rater or several, without the quest for some kind of objective standard of what is, for example, (good) grammar and (good) spelling, and agreement on what importance to attach to particular criteria, there cannot be much reliability.

00-664 Gattullo, Francesca (Università di Bologna / Università di Trieste, Italy; *Email*: rav1905@iperbole.bologna.it). Formative assessment in ELT primary (elementary) classrooms: an Italian case study. *Language Testing* (London, UK), **17**, 2 (2000), 278–88.

This article focuses on research into classroom implementation, in particular on the formative uses of teacher assessment in two primary schools in the Italian state sector where English is taught as a foreign language. In the form of a pilot study—through classroom observation, field notes, teacher and pupil questionnaires, and teacher interviews—the author focuses on procedures used by teachers in the assessment of language learning, the nature of feedback provided for learners and the uses teachers have made of assessment results. In analysing the construct of formative assessment, she borrows the concept of ‘assessment episode’ (Mavrommatis, 1997) and analyses her data according to four phases: (1) evidence collection; (2) decision making against chosen criteria; (3) feedback action; and (4) impact on teaching and learning. On the basis of transcripts of teacher–pupil interaction, she examines emergent processes underlying formative assessment. In a climate where assessment, and related activities, may take up huge amounts of time and energy, this research into teachers’ implementation of assessment represents an empirical contribution and one way of deconstructing assessment processes.

00-665 Hasselgren, Angela (U. of Bergen, Norway; *Email*: Angela.Hasselgren@eng.uib.no). The assessment of the English ability of young learners in Norwegian schools: an innovative

approach. *Language Testing* (London, UK), **17**, 2 (2000), 261–77.

This article describes the development of materials to support classroom based assessment in Norway, a country where there is a deliberate policy not to formally assess (i.e., grade) learners in the early years of foreign language learning. This teacher development strategy aims to upgrade the assessment skills of teachers and to encourage the involvement of primary learners in the assessment of their own progress in language learning. The EVA (EVALuation of English in Schools) Project has systematically developed and trialled materials for teacher use in the assessment of student language learning in Norwegian primary schools. The author identifies characteristics of the younger language learner in Norway which have been influential in the design of the assessment materials, and defines the construct of language ability and the pedagogical principles which underpin the materials developed. Although the analysis of data from national trialling is still in its early stages, the article puts forward some of the positive contributions that this national initiative might make by directly addressing concerns of professional practice and teacher development.

00-666 Ito, Akihiro (Aichi Gakuin U., Japan). Tests as a second language research method: their types, reliability, validity, and variable research results. *ITL Review of Applied Linguistics* (Leuven, Belgium), **127-128** (2000), 1–36.

In second language acquisition (SLA) research various tests are employed for theory building. However, what tests should be administered to participants in a research setting? In order to address the issue, this study was conducted to examine the reliability, validity of tests themselves, and the difference in results according to the difference of test-types, focusing on cross-sectional restrictive relative clause acquisition and relative clause test-types. Among the variety of relative clause tests, four test-types that appear frequently in recent SLA academic publications were selected: Translation, Cloze Procedure, Grammaticality Judgement, and Sentence Combining. The results from 120 Japanese students indicate the following. First, Sentence Combining shows high reliability in internal consistency and the highest validity. Second, the research results change across the different test-types. Third, the concept of ‘test-type (task) related interlanguage variability’ should be explained by the combination of quality-related issues such as ‘measurement error’ and the cognitive demands each test-type requires of the participants. Implications for the practical issues of language test construction for SLA research, educational evaluation, and directions for further research are also discussed.

00-667 Johnstone, Richard (U. of Stirling, UK; *Email*: R.M.Johnstone@stir.ac.uk). Context-sensitive assessment of modern languages in primary (elementary) and early secondary education: Scotland and the European experience.

Language Testing (London, UK), **17**, 2 (2000), 123–43.

Although much more limited in scope and intensity than second language (L2) immersion, MLPS (Modern Languages at Primary School: known elsewhere as FLES, Foreign Languages at Elementary School) is being extensively implemented across the European Union as a reflection of national and European Commission policies. Despite an increase of research interest in the area, issues of MLPS assessment have not been widely addressed. The present author lists good reasons for developing such assessments: gauging the return on a massive public investment; feedback to parents and the public; informing national policy development; and self-evaluation at the level of institutions, departments, teachers and learners. There are, however, major problems at present in doing so across schools: variability of context; embeddedness of children's language in a flow of events; their relative lack of cultural knowledge; unfamiliarity of teachers with concepts of L2 testing; and lack of consensus concerning what MLPS proficiency might mean. The article gives examples from Scotland of how some of these problems were addressed in the pilot phase of the national MLPS initiative of the early to mid-1990s, leading to the development of more comprehensive and standardised assessments in the national Assessment of Achievement Programme intended for the late 1990s and early years of the twenty-first century.

00-668 McKay, Penny (Queensland U. of Technology, Australia). The effectiveness of work samples as an elaboration of profiles. *Babel (AFMLTA)* (North Adelaide, Australia), **34**, 3 (1999), 21–25.

This article discusses the recent profiles for the many foreign languages taught in Australian schools: these have been accompanied by work samples to illustrate the levels, but this approach has limitations. Samples display considerable contextual variations across task types, particularly in the degree of support given and diversity of text-related demands. There is an inevitable tension between the needs to illustrate levels clearly and to provide examples of good practice; samples for higher levels of ability, especially among younger learners, seem scarce. Over-reliance on work samples can lead to a narrowing of the curriculum. The focus on learners' communicative use of the target language to convey meaning marginalises the wider range of desirable learning outcomes. A flexible conception of overlapping levels of progress and ranges of possible scores is considered preferable to a linear sequence of scales. There are also inherent dangers in illustrating and judging a learner's ability through a single task. The range of tasks selected should include an element of unpredictability, minimal teacher input, avoidance of anything limiting learners' performance, promotion of interaction, and topic areas that go beyond personal communication. Teachers should be able to meet to discuss the complexities of profiles and avoid their use in formal testing and reporting.

00-669 McKay, Penny (Queensland U. of Technology, Australia; *Email*: pa.mckay@qut.edu.au). On ESL standards for school-age learners. *Language Testing* (London, UK), **17**, 2 (2000), 185–214.

The recent introduction of literacy benchmarks and assessment regimes in schools which are designed for students with an English-speaking background but which are also used to assess the second language and literacy progress of English as a Second Language (ESL) learners has repercussions for teaching and learning, on professional understandings and on ESL learner self-efficacy. Insights into how to map and monitor ESL language development within the mainstream context have been gained through the development of a number of ESL standards for schools. This article presents principles behind the construction of these standards, drawing on examples of ESL standards developed in Australia, England and Wales, and the United States, and examines how differences in purpose-planning, professional understanding, and reporting-influence how ESL standards might best be constructed. It raises issues of consequential validity which arise in the construction and use of valid ESL standards. Arguments supporting the need for ESL standards in the current climate of English-speaking background-based assessment for all are put forward through a close analysis and understanding of the construction and impact of different types of ESL standards.

00-670 Meara, Paul and Fitzpatrick, Tess (U. of Wales Swansea, UK; *Email*: p.m.meara@swan.ac.uk). Lex30: an improved method of assessing productive vocabulary in an L2. *System* (Oxford, UK), **28**, 1 (2000), 19–30.

This paper describes a test of productive vocabulary. The authors believe it easy to administer and suitable for making straightforward assessments of the productive vocabulary of non-native speakers of English. The paper first describes the methodology developed by the authors, then shows how this methodology could be used to make interesting comparisons between productive and receptive vocabulary in second language (L2) learners. The test requires students to produce a set of word association responses to a small set of stimulus words. The stimulus words are chosen so that with native speaker students they typically generate a wide range of responses, and a high proportion of low frequency responses. The paper argues that a test of this sort might tap the extent of non-native speakers' productive vocabulary more effectively than some other tests in current use.

00-671 Norton, Bonny and Stewart, Gail (U. of British Columbia, Canada). Accountability in language assessment of adult immigrants in Canada. *The Canadian Modern Language Review / La Revue canadienne des langues vivantes* (Toronto, Ont.), **56**, 2 (1999), 223–44.



This article addresses the challenges the authors faced in developing a task-based language assessment instrument for adult immigrants in Canada which is accountable to the needs of diverse groups of stakeholders, including learners, teachers and administrators. The test, the Canadian Language Benchmarks Assessment (CLBA), is designed to place adult newcomers in language programmes appropriate for their level of proficiency in English and to assess progress in these programmes. The authors note that, while stakeholders wanted the assessment tasks to be authentic and realistic, many were concerned that authentic tasks are culturally biased. These concerns were associated with stakeholders' theories of language, their conceptions of test bias, and their understanding of the purpose and use of the test. The authors suggest that when the low-stakes CLBA is used for its intended purpose, its results can satisfy the imperative for accountability in the context of large-scale, system-wide assessment.

00-672 Nunn, Roger (U. of Kochi, Japan; *Email*: nunn@cc-kochi-u.cc.jp). Designing rating scales for small-group interaction. *ELT Journal* (Oxford, UK), **54**, 2 (2000), 169–78.

Classroom activities in small groups provide opportunities for practising important interaction skills such as distributing and competing for opportunities to speak, holding the floor, adjusting to the contributions of other speakers, and negotiating real understanding when exchanging information, opinions, feelings, and attitudes. A rating scale is proposed here as a practical means of addressing the difficult task of assessing both the level of a particular communicative performance in a small group and the general ability to perform in small-group conversations over time. This paper argues that theoretical difficulties of designing and using rating scales for this purpose, while requiring serious consideration, are outweighed by practical advantages. It is argued that rating scales not only report test performances: they can also guide the teaching process, defining the principles for the construction of both assessment and classroom tasks and providing teachers (and students) with achievable goals which they themselves have formulated in writing.

00-673 Qian, David D. (Educational Testing Service, Princeton, USA). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review / La Revue canadienne des langues vivantes* (Toronto, Ont.), **56**, 2 (1999), 282–307.

The empirical study reported here explored the relationships between depth and breadth of vocabulary knowledge and reading comprehension in English as a Second Language (ESL). Using multivariate analyses, the study examined the roles of depth and breadth of vocabulary knowledge in assessing the performance of a group of young adult ESL learners with a minimum vocabulary size of 3,000 word families in carrying out

general academic reading comprehension tasks. The results support the hypotheses that (1) scores on vocabulary size, depth of vocabulary knowledge, and reading comprehension are highly, and positively, correlated; and (2) scores on depth of vocabulary knowledge can make a unique contribution to the prediction of reading comprehension levels, in addition to the prediction afforded by vocabulary size scores. It is suggested that the findings from this study call for a recognition of the importance of improving depth of vocabulary knowledge in learners' ESL learning processes.

00-674 Rea-Dickins, Pauline (U. of Bristol, UK; *Email*: P.Rea-Dickins@bristol.ac.uk) and **Gardner, Sheena**. Snares and silver bullets: disentangling the construct of formative assessment. *Language Testing* (London, UK), **17**, 2 (2000), 215–43.

This article explores the nature of formative assessment in a primary (elementary) language learning context. The research is situated in nine inner-city schools in England where an Early Years Intervention Project is being implemented to address problems of low levels of achievement in English, with specific reference to the language support of learners for whom English is an Additional Language. School-based assessment data are presented and analysed in relation to the construct of formative assessment. It is argued that the distinctions between formative and summative assessment are not as straightforward as sometimes portrayed, and that the interplay between reliability and validity for purposes of class-based assessment is highly complex.

00-675 Teasdale, Alex and Leung, Constant (King's Coll., London, UK; *Email*: alex@teasdales.freemove.co.uk). Teacher assessment and psychometric theory: a case of paradigm crossing? *Language Testing* (London, UK), **17**, 2 (2000), 163–84.

Alternative assessment (as opposed to formal testing) is gaining a great deal of attention in current educational discussion. This article attempts to address some of the epistemic and practical issues facing alternative assessment, with particular reference to teacher assessment of spoken English as an Additional Language / Second Language (EAL/ESL) in the early years of primary (elementary) education in England and Wales. The authors first examine the claims made by advocates of alternative assessment in terms of validity and educational relevance. It is argued that such claims are founded on an uneasy articulation of different principles underpinning psychometric measurement and pedagogy. They next look at some of the reasons why psychometric approaches may not provide an adequate response to pedagogic and policy developments. They then discuss some of the theoretical and practical problems involved in teacher assessment of speaking, focusing on learners with EAL in primary education. They focus on the importance of clarity about the epistemological bases of different types of assessment. Additionally, the article highlights the need to be alert

to the ways in which political and ideological concerns—together with the influence of professional (teaching) culture—are influential in shaping the properties of assessment systems.

00-676 Zangl, Renate (Karl-Franzens U., Graz, Austria). Monitoring language skills in Austrian primary (elementary) schools: a case study. *Language Testing* (London, UK), **17**, 2 (2000), 250–60.

This account of work-in-progress represents a contribution at the interface of programme evaluation and second language acquisition. The article describes the assessment procedures used in two different foreign language learning programmes at primary school level in Austria: the Vienna Bilingual Schooling and the Lollipop Programme. In the first of these, which takes a longitudinal approach, the researcher traces the development of spontaneous speech and the linguistic systems—morphology, syntax, and semantics/lexicon—in language learners from the first to fourth grade (ages 6–10). Some key principles which it is considered should frame approaches to the assessment of young language learners are also raised.

Teacher education

00-677 Allen, Linda Quinn (U. of Missouri-St. Louis, USA). Culture and the ethnographic interview in foreign language teacher development. *Foreign Language Annals* (New York, USA), **33**, 1 (2000), 51–7.

This article starts from the assumption that, in order to design and implement foreign language (FL) instruction which integrates culture with language, teachers must first have a deep understanding of the target culture and an awareness of their own culturally conditioned and individually formed beliefs, attitudes, and values. The article presents the ethnographic interview as a strategy in FL teacher development. Excerpts from preservice teachers' reports on an ethnographic interview assignment are seen as providing evidence that the process of conducting the interview facilitates the development of the requisite understanding and awareness. Given the experience with the ethnographic interview, it seems more likely that the preservice teachers' new understandings will influence the instruction they provide to their own FL students; and that culture instruction may thus cease to be an add-on and move towards becoming a more integrated part of FL classroom practice.

00-678 Amiri, Faramarz (U. of Wolverhampton, UK; *Email*: f.amiri@wlv.ac.uk) IT-literacy for language teachers: should it include computer programming? *System* (Oxford, UK), **28**, 1 (2000), 77–84.

This article examines some of the issues related to the information technology (IT) training of the language

teachers in MA TEFL/TESOL courses and argues that teachers at this level should be trained not merely as consumers but as both *consumers* and *producers* of computer-based materials. The article focuses on the role of computer programming in MA courses and concludes that knowledge of programming is important for enabling language teachers to get involved in the design and development of computer-based materials and should be included as a part of the IT training for language teachers. It is concluded that there is a need for more specialised MA courses for language teaching that concentrate on various aspects of IT, including end-user programming, learning, instructional design and second language acquisition theories.

00-679 Crandall, JoAnn (U. of Maryland, Baltimore County, USA). Language teacher education. *Annual Review of Applied Linguistics* (New York, USA), **20** (2000), 34–55.

This paper reviews recent work on language teacher education programmes and identifies four major areas of concern. First, there is a move from product-oriented theories to process-oriented; second, there is dissatisfaction with the way such programmes have prepared teachers for the realities of the classroom; third, there is growing recognition of the importance of teachers' prior learning experiences in shaping effective teaching; and finally, there is growing concern for the role of teachers as active contributors to developing theory and directing their own professional development. Particular attention is paid to the shift from methods to methodology and the need to examine underlying teacher beliefs and thinking to help balance more product-oriented conceptions of language teaching. The final part of the paper examines the role of native and non-native teachers and concludes that it is no longer possible to provide an acceptable description of the native speaker. Furthermore, the link between native-speaking proficiency and professional competence is often misconstrued, to the detriment of the non-native speaking teacher. The article is followed by both an annotated and unannotated bibliography.

00-680 Hayes, David (Primary English Language project, Sri Lanka; *Email*: davidh@lanka.ccom.lk). Cascade training and teachers' professional development. *ELT Journal* (Oxford, UK), **54**, 2 (2000), 135–45.

In their concern with 'improved' curricula and 'more effective' teaching-learning methods, education ministries often use the 'cascade' model to attempt to effect large-scale change at the classroom level. Experience of cascades in in-service development has tended to show, however, that the cascade is more often reduced to a trickle by the time it reaches the classroom teacher, on whom the success of curricular change depends. This paper examines the experience of a nationwide in-service teacher development project in Sri Lanka which aims to remedy the potential deficiencies of cascade models of teacher development. It shows how project